An Extraction Method for Glacial Lakes Based on Landsat-8 Imagery Using an Improved U-Net Network

Yi He, Sheng Yao, Wang Yang, Haowen Yan, Lifeng Zhang, Zhiqing Wen, Yali Zhang, and Tao Liu

***Abstract*—Remote sensing monitoring of glacial lakes is an indispensable tool for identifying and preventing glacial lake disasters. At present, the existing extraction methods of glacial lakes based on Landsat remote sensing image have achieved remarkable results, but the algorithms used lack the ability to analyze glacial lake spectral and shape and texture features, and require manual design parameters to fine tune the automation of the algorithm. As a result, it cannot mine the depth features of glacier lakes in remote sensing images accurately enough. To address these challenges, this study designed a self-attention mechanism module U-net network that enhances the propagation of features, reduces information loss, strengthens the weight of glacial lake areas, restrains the weight of irrelevant features, reduces the influence of low image contrast on the model, and deals with the variety of pixel categories in glacial lakes. These features improve the performance of the model. Based on Landsat-8 images, we first extracted glacial lakes in large-scale alpine areas using a U-net network model. To make it a self-attention U-net network, we introduced the attention mechanism into the step connection part of the U-net network to adjust feature weight, focus on learning glacial lake features, and strengthen the network to extract the glacial lake features. Finally, we selected the combination of bands 3, 5, and 6 and all bands of Landsat-8 images sing the self-attention U-net network to extract glacial lakes in the study area and compared and analyzed the extraction results. The experimental results and analyses revealed that the proposed method can effectively segment glacial lakes from Landsat-8 remote sensing images. Its effectiveness was proven by different evaluation indices. Compared with a standard U-net network, the true positive for the combination of 3, 5, and 6 bands increased by 15.95% and for all bands by 5.79%. The area under curve for the whole study area reached 85.03% for all bands. The improved U-net network can, thus, meet the real time needs of glacial lake disaster information acquisition.**

Manuscript received April 2, 2021; revised May 12, 2021; accepted May 26, 2021. Date of publication June 1, 2021; date of current version July 14, 2021. This work was supported in part by the National Scientific Foundation of China under Grants 41761088, 41371435, 41761014, 41761082, and 41861059, in part by LZJTU EP under Grant 201806, in part by the Natural Science Foundation of Gansu Province under Grant 20JR10RA249, in part by the Youth Science and Technology Foundation of Gansu Province under Grant 20JR10RA272, in part by the Tianyou Youth Talent Lift Program of Lanzhou Jiaotong University, in part by the Department of Education of Gansu Province under Grant 2019A-043, and in part by China Postdoctoral Science Foundation under Grant 2019M660092XB. (*Corresponding author: Haowen Yan.*)

The authors are with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China, with the National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China, and also with the Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China (e-mail: heyi@mail.lzjtu.cn; 1024685056@qq.com; yyangwang48@gmail.com; 947258095@qq.com; 764324437@qq.com; 936784011@qq.com; 493497605@qq.com; 25951804@qq.com).

Digital Object Identifier 10.1109/JSTARS.2021.3085397

***Index Terms*—Attention mechanism module, glacial lake, multispectral image, remote sensing, U-net.**

I. Introduction

GLACIAL lakes are natural bodies of water formed by modern glacial meltwater as the main supply source of stagnant water in moraine ridge depressions [1]. Glacial lakes are the incubators of alpine glacial disasters and are important in the study of mountain disasters [2]. Many glacial lake collapse events have occurred in recent years, resulting in significant losses in life and property [3]. The real-time monitoring of glacial lakes is, therefore, essential. Glacial lakes are a special kind of lake with diverse features found at high altitudes. The geographic information contained in optical remote sensing images is complex, and the sensors of optical remote sensing satellites have multiband data, which are affected by different topographies and also carry complex semantic information [4]. In remote sensing images, glacial lakes can appear similar to some ground features (such as mountain shadows and melting glaciers), so it is difficult to extract the large-scale remote sensing data on glacial lakes accurately.

At present, remote sensing extraction methods of glacial lake boundary information mainly include manual digitization, spectral information, traditional machine learning, and image semantic segmentation. The manual digitization method is highly accurate but time- and labor-consuming, as well as inefficient, so it is suitable only for small-scale glacial lake information extraction [5], [6]. The representative algorithms of the spectral information method include threshold and water body index methods, which are usually used together. The idea is to use the normalized water body index [normalized difference water index (NDWI)] to set the global threshold, eliminate nonglacial lake information, and extract the glacial lake information. The spectral information method is simple and effective, but it is suitable only for small areas because the spectral characteristics of glacial lakes and some ground objects (such as mountain shadows and melting glaciers) in large areas are similar [7].

The representative algorithms of the traditional machine learning include decision trees and neural networks. Decision tree and neural network methods are suitable for areas with prior knowledge, which limits their universal application [8]–[13]. Image semantic segmentation methods include object-oriented segmentation and distributed iteration. The image semantic segmentation method can accurately extract glacial lake boundaries by using image features, but it requires postprocessing operations, and the segmentation scale of the area where glacial lakes are located and the size of glacial lakes vary, which also limits the wide application of the method [14], [15]. The existing glacial lake extraction algorithms generally lack the ability to analyze the spectrum, shape, texture, and other features of glacial lakes because of the differences in pixel spectra between glacial lakes, which makes it difficult to extract glacial lake data on a large scale. The algorithms also require more manual design parameters to realize their automation. Deep learning is an emerging multilayer neural network learning algorithm that can extract intrinsic and deep features [16]. Without manual intervention, advanced features can be obtained from original features to improve the accuracy of classification [17]–[19]. Deep learning algorithms have been widely used in the study of remote sensing image classification and have achieved remarkable results [17], [19], [20]–[23]. Compared with the conventional methods, deep learning can better deal with the complex features of remote sensing images and has stronger autonomous learning ability [24].

Deep full convolutional neural (FCN) networks show advantages in the field of image semantic segmentation and uses its code-decoding structure to segment ground objects and achieves good results [25], [26]. On the basis of FCN, Wu *et al.* [27] proposed a U-shaped symmetric network (U-Net), which can integrate low- and high-dimensional features, which greatly improves segmentation accuracy. U-net networks have been widely applied to the classification of remote sensing images [28], [29]. Later, scholars extracted water bodies using a U-net network and achieved good results, but it underperformed when segmentation targets were very small [30], [31]. When lake and mountain areas account for only a small fraction of the cutting area, the traditional networks run into a limitation, which is that the U-net skip connection directly to the deep features of shallow features of encoder and decoder combining prone to semantic gap [32]. This study used jump connection module replacement parts for the space attention mechanism. This way, the jump part of the network can focus on the input feature subset and select specific and noteworthy input, thus solving the problem of overloading original information and the semantic gap problem of the U-net jump part due to the combination of features of different dimensions. Thus, it improves the classification accuracy of small targets. We define the network designed in this study as a self-attention U-net.

In this study, 11 bands of Landsat-8 were used to extract a glacial lake. NDWI and normalized differential snow index (NDSI) are standard methods of extracting water. The 3, 5, and 6 bands of Landsat-8 image are sensitive to water, so this experiment also tried to combine these bands to extract the glacial lake information. The objectives of this study are as follows.

1) Planning a U-net network to extract large-scale glacial lake information from the combination of 3, 5, and 6 bands and all bands of Landsat-8 remote sensing images.
2) Developing a self-attention U-net network framework to improve glacial lake extraction performance and accuracy.
3) Extracting glacial lake information from Landsat-8 images in 2018 based on the self-attention U-net network.

This study provides a supporting method for rapid and intelligent monitoring of glacial lake information along the Sichuan–Tibet railway.

The rest of this article is organized as follows. Section II describes the study area. Section III describes the U-net and self-attention U-net network model. Section IV presents the results and analysis. Finally, Section V concludes this article.

II. Study Area

The Alatau mountains of Tianshan was selected as the study area. Alatau mountain belongs to the Tianshan mountain system, which is located in the north of the Bortala Mongol autonomous prefecture in Xinjiang (79°30′–81°45′E, 44°40′–45°20′N). The mountain runs from east to west, and its northern slope is in the Republic of Kazakhstan. The snow line is low at 3500 m. At 3800 m, there are mainly suspended glacial and bucket glacial lakes, most of which are distributed in the north. The end drops to 3300 m above sea level, the ridge elevation is 4000 m, and the highest peak is 4570 m. The annual precipitation in the alpine belt can reach 1000 mm. The north side, in Kazakhstan, is moist, with the snow line below 3600 m. The south side, in China, is dry, and the snow line can rise to 3900 m. About two-third of glacial lakes are distributed in the north side of Kazakhstan. In recent years, many small glacial lakes have developed of various types and with rich characteristics [2]. The study area is shown in Fig. 1.

III. Methodology

A number of Landsat-8 images were first selected and embedded into the study area. Then, each band image was cut into chunks to construct a dataset. Then, U-net and improved U-net (self-attention U-net) networks were used to extract glacial lakes, which are depicted in Fig. 2. Kappa coefficient, F1-score, Mean Intersection over Union (MIoU), and area under curve (AUC) were used for evaluation. The following sections provide more details on the individual aspects of the methodology.

A. Data Source

This research selects Landsat-8 satellite remote sensing data (orbit numbers 147 028, 147 029, and 148 029) in 2018 (https://www.usgs.gov). Landsat-8 data include 11 bands, and the specific sensor parameters are given in Table I. First, the remote sensing image is preprocessed in ENVI software, including atmospheric correction, geometric correction, and radiation calibration, then the band layers are superimposed together, and the multiscene image is embedded to make it become the whole multispectral image of the study area. All 11 bands were selected for the experiment in this study. The spatial resolution of two bands in Landsat-8 TIRS is 100 m, and we resampled them to the same resolution as the other bands (30 m). Normalized differential water index (NDWI) and (NDSI are classic methods to extract water. The 3, 5, and 6 bands used are sensitive to water, so this experiment also tries to combine 3, 5, and 6 bands

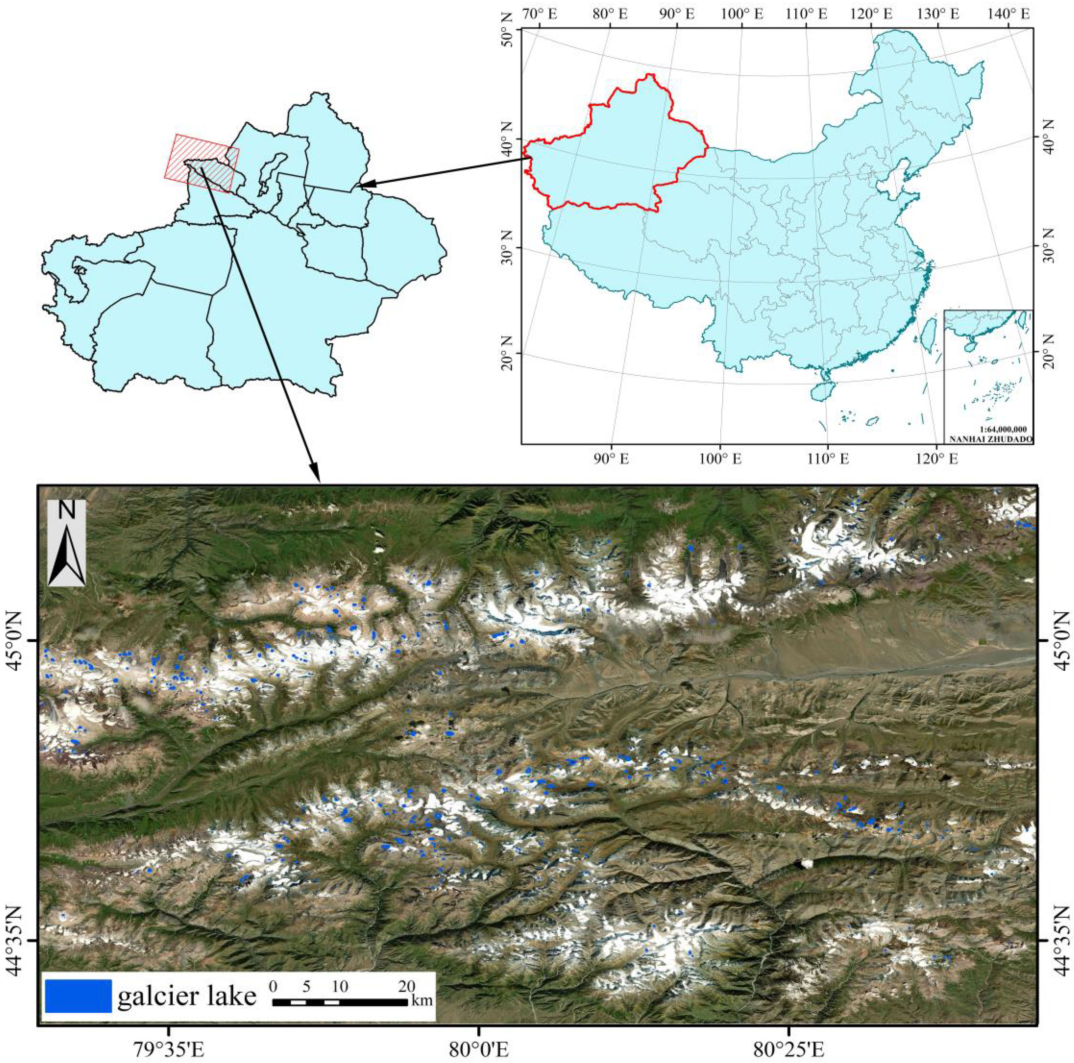


Fig. 1. Location of the study area.

TABLE I
LANDSAT-8 SATELLITE SENSOR PARAMETERS

Sensor	Band	Wavelength rangen (um)	Signal noise ratio (SNR)	Spatial resolution (m)
OLI	1 Coastal	0.43-0.45	130	30
	2 Blue	0.45-0.51	130	30
	3 Green	0.53-0.59	100	30
	4 Red	0.64-0.67	90	30
	5 NIR	0.85-0.88	90	30
	6 SWIR1	1.57-1.65	100	30
	7 SWIR2	2.11-2.29	100	30
	8 Pan	0.50-0.68	80	15
	9 Cirrus	1.36-1.38	50	30
TIRS	10 TIRS1	10.60-11.19	0.4K	100
	11 TIRS2	11.50-12.51	0.4K	100

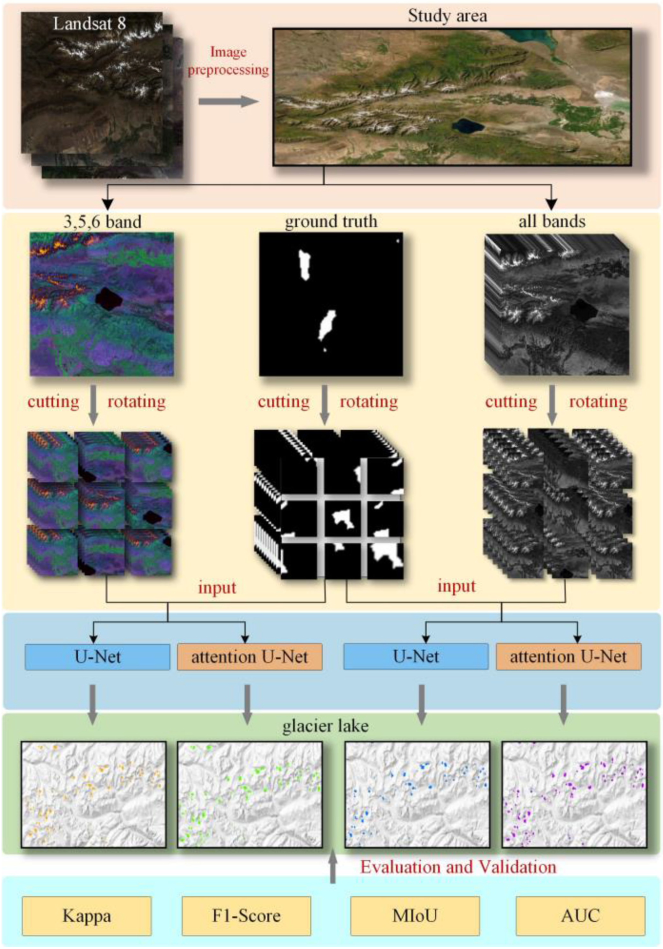


Fig. 2. Glacial lake extraction flowchart.

to extract the information of the glacial lake to improve the efficiency.

B. Dataset Construction

1) Ground Truth: The accurate glacial lake boundary information in the study area was extracted by visual interpretation of Google Earth images and was converted into a binary image as the ground truth for samples in which the foreground (white) represents glacial lake areas and the background (black) represents nonglacial lake areas (as shown in Fig. 3).

2) Training Data: Remote sensing images (the combination of 3, 5, and 6 bands and all bands) were cut and enlarged. The cutting method defines three scaling ratios, which are 1, 2, 4, and the output image size is 256 × 256, 128 × 128, and 64 × 64. In the 256 × 256 size cutting, we adopted the overlapping cutting strategy with a step size of 128. Therefore, most areas with glacial lakes were trained 2–4 times in the neural network, and they were all at different positions of the image after cutting. The overlapping strategy was not adopted in the cutting of 128 × 128 and 64 × 64 sizes because when the dataset was resized to 256×256, the distortion of too much scaled data would affect the precision of network training. The augmentation method mirrors and rotates (rot 90°) the original image. The output image should contain not only the original glacial image but also the image that mirrors and rotates (rot 90°) the original image in order to enrich the training data. The same operation is carried out on the ground truth to get a data pair with the processed images. In order to simplify the training dataset, the images without glacial image is excluded from the dataset.

3) Dataset: The training image and ground truth were read separately according to the following path: each time, the image and the corresponding label were scaled to 256×256 and normalized; the proportion of training set:test set:verification set was set to 8:1:1, uniform sampling. Finally, 2553 training samples, 320 verification samples, and 320 test samples were generated. Fig. 3 shows some representative samples in the dataset formed after sample segmentation.

C. U-Net Network

A U-net network is a U-shaped network structure, which can obtain context and location information at the same time. Its design was originally intended to solve the problem of medical image semantic segmentation [33]. U-net is a more refined design based on the basic structure of an FCN network and it is also more efficient, replacing the optimized FCN networks. It is mainly composed of an encoder, decoder, and skip connection [26], [33], [34]. The encoder is used to extract spatial features from images and reduce spatial dimensions. It divides the feature map into five scales, each of which contains two convolution layers with the same number of output channels and 3×3 convolution kernel size and is connected by a maximum pool with a step size of 2. Through scale-by-scale convolution sampling, features of different dimensions are extracted and retained from the context of the feature map. It is composed of two 3×3 convolution layers (ReLU) and 2×2 max polling layers (stride is 2). After each downsampling, the number of channels is doubled, and the last two 3×3 convolution operations connect the encoder to the decoder. The convolution layer is used to extract image features, and the lower sampling layer is used to filter unimportant high-frequency information, reduce the feature dimensions, and increase the receptive field. Repeated convolution and pooling operations can fully extract the high-level features of the remote sensing image. The decoder is used to construct a segmentation map according to the encoder features to gradually restore the details and spatial dimensions of the remote sensing image. In the decoder, the scale division of the feature map is partially symmetrical with that of the encoder, and the feature extraction part of each scale is composed of two 3×3 convolution layers with the same number of output channels. Each scale is also connected to an upper sampling layer of size 2 and a convolution layer with a convolution kernel size of 2×2. Finally, the 1×1 convolution layer returns to the pixel classification, and the input of each decoding block is fused with the output of the corresponding layer coding block as the input of the next deconvolution layer, so as to reduce the information loss caused by downsampling in the coding block. The problem of glacial lake extraction belongs to two categories, but there is only one output channel. Skip connection transfers the output from encoder to decoder in series with the output of the upsampling

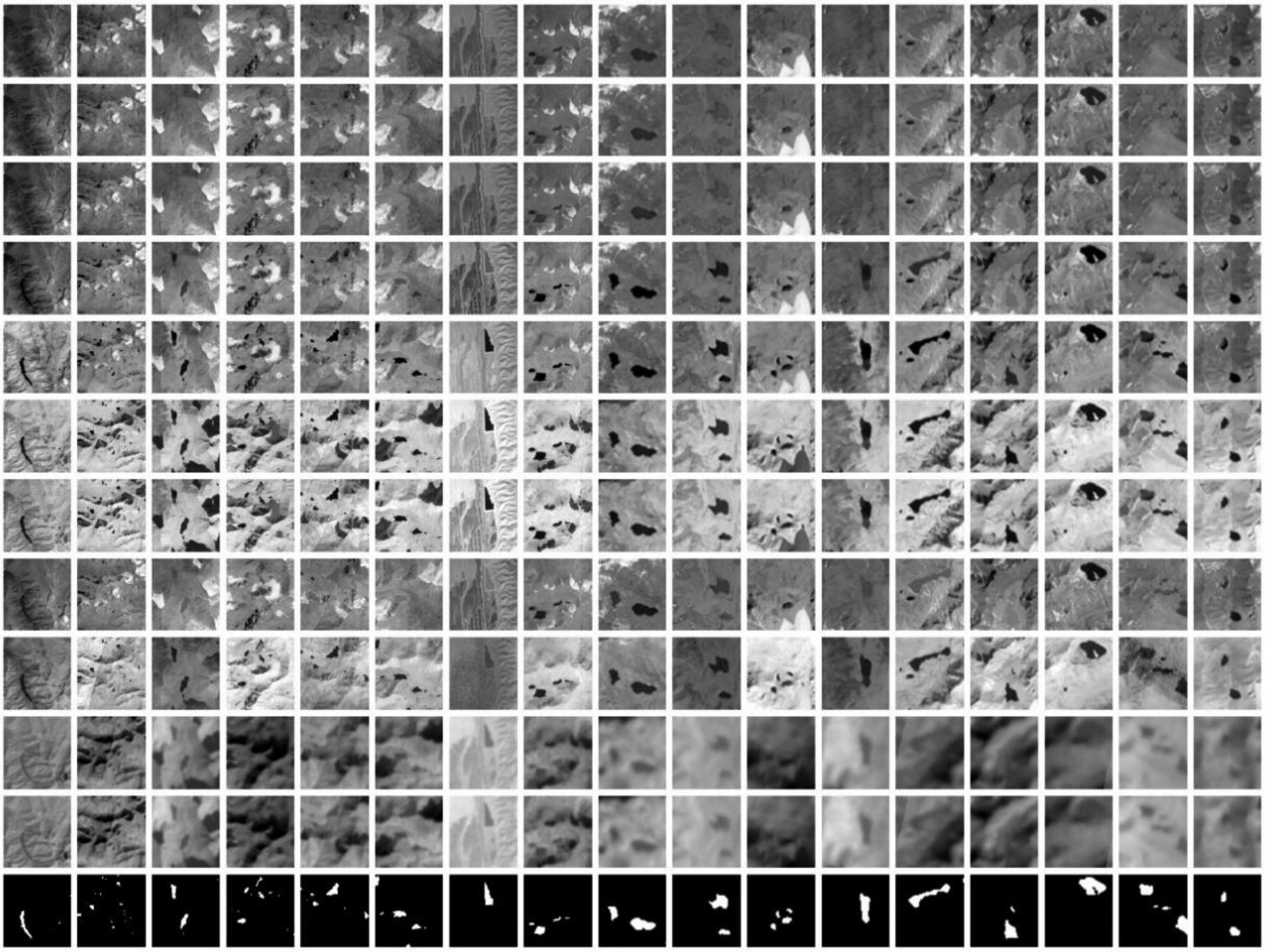


Fig. 3. Sample of the partial dataset after sample segmentation (all bands, ground truth).

operation and propagates the cascade feature map to subsequent layers to retain as much detail as possible and improve the resolution and edge accuracy of the final segmentation result. Padding of all convolutional layers is set to the size of output feature graph consistent with the input, that is, the "same" mode is used, which has little impact on segmentation accuracy in the classification task, and is easy to make datasets and handle subsequent parts. In this study, U-net is used as the basic network for the glacial lake boundary information extraction model, the structure of which is shown in Fig. 4.

D. U-Net Network Based on Self-Attention Mechanism

Remote sensing image data contain various kinds of complex feature information, such as glaciers, vegetation, bare land, and mountain shadows. In this study, only glacial lake features are extracted, and other features are used as background processing. The complex background greatly interferes with the accuracy of glacial lake extraction [35]. The attention mechanism is designed to selectively ignore a part of the information to carry out the weighted aggregate calculation of the rest of the information. Its basic function is to highlight a core part of the feature map to become the input of the attention feature so that the model pays more attention to the relevant information [36]. Therefore, this study proposes to introduce the attention mechanism into the step connection part of the U-net network to adjust the feature weight, focus on learning glacial lake features, and strengthen the network for extracting glacial lake features.

A structural diagram of the attention gate (AG) is shown in Fig. 5, where g is the feature map matrix of the decoding part and x is the feature map matrix of the coding part. The self-attention U-net network structure is shown in Fig. 6. The encoder and decoder parts use the same scale, which is connected to the AG module to realize the attention mechanism. First, the AG module carries out a 3×3 convolution operation by combining the feature map of the same scale as the encoder part with the result of the upper sampling from the decoder part, with the number of output channels being half that of the scale encoder and decoder parts, so as to extract the fused features on a coarse scale while eliminating the irrelevant noise. ReLU has an activation function to suppress overfitting from the AG module [37]. Then, the convolution layer with 1 output channel and 1×1 convolution kernel size is used to output the attention weight matrix, and a sigmoid function is selected as the activation function to output the normalized weight index. Sigmoid functions are used as a regression classification method for dichotomy problems [28].

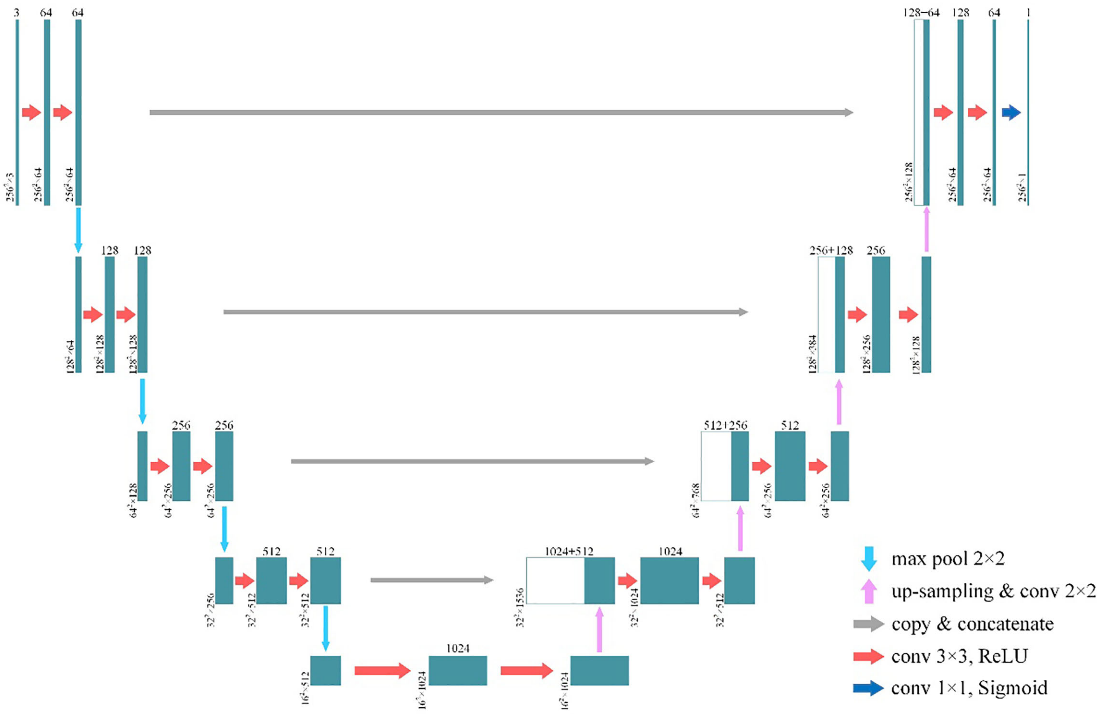


Fig. 4. U-net network structure.

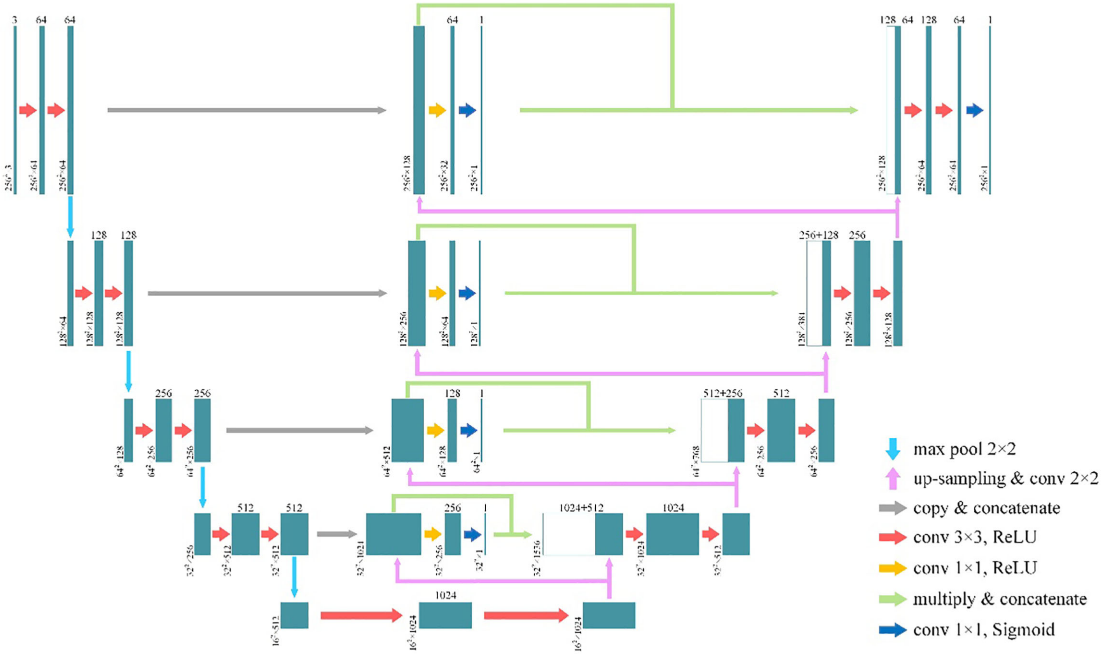


Fig. 5. Self-attention mechanism model.

By multiplying the normalized attention weight matrix and the fusion result of the first step of the AG module, the feature fusion result with attention weight is obtained. Finally, the result is fused with the result of the 2×2 convolution after partial upsampling from the decoder. The AG mechanism not only satisfies the multiscale feature fusion in semantic segmentation but also solves the traditional U-net deficiency of fixed weights in the same scale skip connection. Padding of all convolutional layers is set to the "same" model, that is, the size of the output feature map is set to be consistent with the input.

The specific operations of the self-attention mechanism are as follows.

Feature weight extraction:

$$W_{\mathrm{concat}(x,g)} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} f(i,j) \tag{1}$$

where g is the feature map matrix of the decoding part, x is the feature map matrix of the encoding part, H and W represent the length and width of the feature map, respectively, $W_{\mathrm{concat}(x,g)}$ is

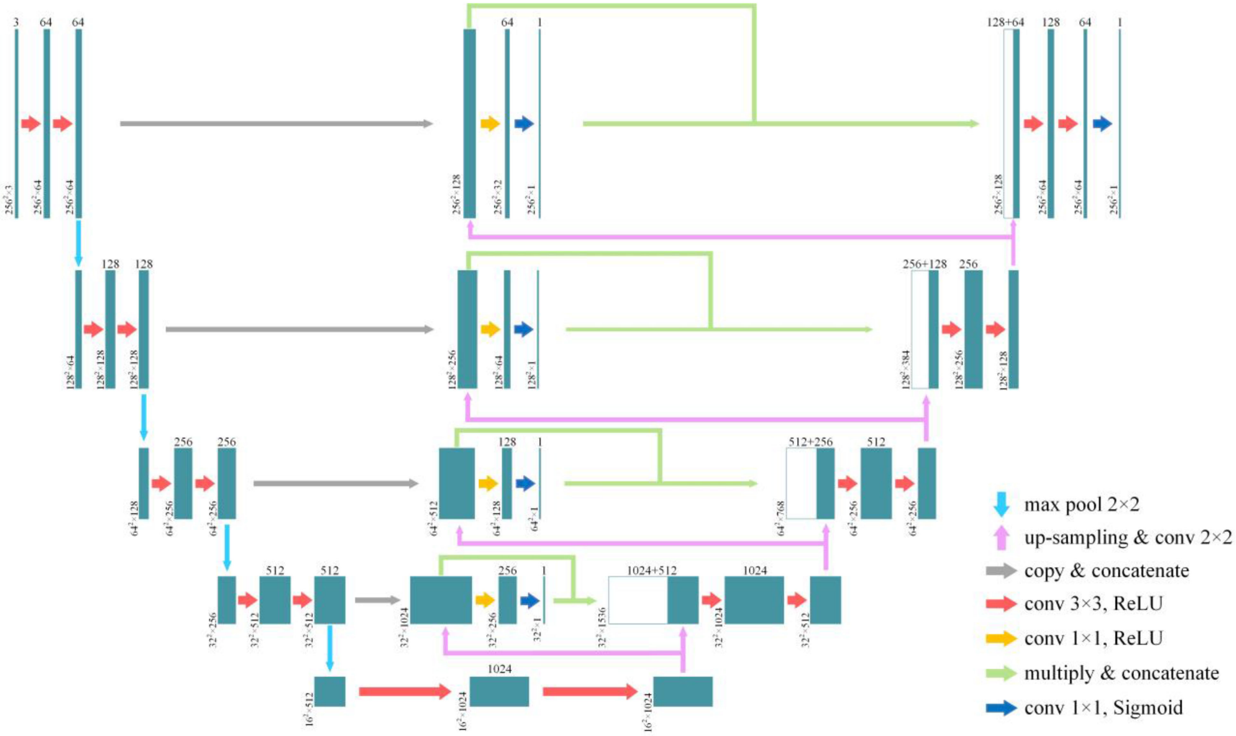


Fig. 6. U-net network structure of attention mechanism.

TABLE II
BASIC SYSTEM PLATFORM CONFIGURATION

Project	Operating system	CPU	Memory	Hard disk	GPU
Content	Microsoft Windows 10 Professional Workstation version 2004	AMD Ryzen 9 4900HS with Radeon Graphics	16G 3200MHz	Intel Q660 1TB	NVIDIA GeForce RTX 2060 with Max-Q Design GDDR6 @ 6GB (192 bits)

the feature weight matrix, and *i* and *j* correspond to the position of pixels in the feature map.

By combining the feature map *x* of the encoding part and the feature map *g* of the decoding part by formula (1), the weight matrix of the feature map $W_{\text{concat}(x,g)}$ is obtained.

Feature weight update:

$$q_{\text{att}} = \text{ReLU}\left(W_{\text{concat}(x,g)}\right) \tag{2}$$

$$\alpha = \text{Sigmoid}(q_{\text{att}}\left(W_{\text{concat}(x,g)}, \Theta_{\text{att}}\right) \tag{3}$$

where ReLU is the activation function. Θ_{att} is a set of parameter linear transformations and bias terms.

E. Training Model

FCN network computing is calculation intensive and consumes a lot of video memory during training, which requires high-level hardware. If limited by the glacial and experimental environment, it will pursue a balance in the platform. The model of this article is based on the deep learning framework Keras. The deep learning experimental environment is built according to the current mainstream configuration environment. The basic configuration is given in Table II.

The important software configuration for this study is given in Table III. According to the operation of each model, the reference can install the corresponding software package appropriately to speed up the operation of the model [38].

When the batch size is set to 8, the convergence rate of a large batch size is lower than that of a small batch size with the same limited capacity for computation. A too-large batch size may make the neural network nonconvex, and there may be multiple local optimal or saddle points, that is, the eigenvalues of the corresponding Hesse matrix are both positive and negative. Therefore, in practical engineering, a small batch of sample set minibatches is optimal from the point of view of convergence speed, and the batch size often varies from tens to hundreds but generally no more than a thousand. Here, due to the hardware limitations of video memory, the batch size is set to 8.

TABLE III
IMPORTANT SOFTWARE CONFIGURATION

Project	Graphic driver	CUDA version	Python	Keras	Tensorflow
Content	456.43	V10.1.105	3.7.6	2.4.0	2.3.0

TABLE IV
ACCURACY EVALUATION CONFUSION MATRIX

Confusion Matrix		Prediction	
		True	False
Ground truth	True	True Positive TP	False Negative FN
	False	False Positive FP	True Negative TN

Considering the computational efficiency of the model, the hardware capacity, and the desired accuracy of the results, the number of iterations over the course of the experiment was set to 128. Adaptive moment estimation (“Adam”) was selected as the optimizer [39]. The Adam algorithm is equivalent to a combination of the RMSProp gradient descent and momentum gradient descent methods. It has a high convergence speed and good learning effect and is suitable for all kinds of neural networks. The learning rate was set to 10^{-4}. The binary cross entropy (BCE) is selected as the loss function. The target of the neural network in this article is the binary classification problem. When the BCE is used as the loss function, the gradient of the final output layer has nothing to do with the derivative of the activation function but is only proportional to the difference between the output value and the real value. Therefore, when the model approaches the real value, the gradient remains at a high state and the convergence speed of the model remains fast. After iterative training over the whole study area, the network finally converged.

The BCE loss function of the second classification arrived as follows. There are only positive and negative examples in the second classification, and the probability sum of the two is 1, so there is no need to predict a vector; therefore, only one probability is needed. The definition of the loss function is as follows:

$$\mathrm{BCE}(x)_i = -\sum_{i=0}^{n} \left(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \right) / 2 \quad (4)$$

where *y* is the probability that the model predicts that the sample is a positive example, and *y* is the sample label; if the sample belongs to a positive example, the value is 1, otherwise the value is 0.

F. Evaluation Index

To quantitatively evaluate the performance of the glacial lake boundary information extraction model, the following indicators were used to evaluate its accuracy: kappa coefficient, precision rate (*P*), recall rate (*R*), F1-score (F1), MIoU, and AUC. When comparing the extraction results, the following four pixel evaluation categories were considered, and several indicators were calculated directly from the confusion matrix, as given in Table IV.

Kappa coefficient:

$$\mathrm{Kappa} = \frac{N \sum_{i=1}^{n} X_{ii} - \sum_{i=1}^{n} (X_{i+} X_{+i})}{N^2 - \sum_{i=1}^{n} (X_{i+} X_{+i})} \quad (5)$$

where *n* is the total number of columns in the confusion matrix (the total number of categories); X_{ii} is the number of samples in the *i* row and *i* column of the confusion matrix, that is, the number of samples correctly classified; X_{i+} and X_{+i} are the total number of samples in row *i* and column *i*, respectively; and *N* is the total number of samples used for accuracy evaluation.

Precision rate (*P*):

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \quad (6)$$

Recall rate (*R*):

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \quad (7)$$

F1-score (F1) is the harmonic average of precision *P* and recall *R*:

$$\mathrm{F1} = 2 \times \frac{P \times R}{P + R}. \quad (8)$$

MIoU for the binary classification problem:

$$\mathrm{MIoU} = \frac{\frac{\mathrm{TP}}{\mathrm{FP+TP+FN}} + \frac{\mathrm{TN}}{\mathrm{FP+TN+FN}}}{2}. \quad (9)$$

AUC:

$$\mathrm{AUC} = \frac{\sum_{\mathrm{ins}_i \in \mathrm{positiveclass}} \mathrm{rank}_{\mathrm{ins}_i} - \frac{M \times (M+1)}{2}}{M \times N} \quad (10)$$

where $\mathrm{rank}_{\mathrm{ins}_i}$ represents the serial number of the sample in article *i*, that is, the probability score is ranked from small to large, ranking in the rank. *M* and *N* are the numbers of positive samples and negative samples, respectively. $\sum_{\text{ins}_i \in \text{positiveclass}} \text{rank}_{\text{ins}_i}$ means that only the rank sequence numbers of the positive samples are added.

IV. Results and Analysis

In this study, U-net and self-attention U-net network models were used to extract a large-scale glacial lake in the Alatau mountains based on a Landsat-8 remote sensing image. The extraction results are shown in Fig. 7(a).

A. Accuracy Analysis of Glacial Lake Extraction Results

From the comparison and analysis of the segmentation result map and ground truth in subjective evaluation (see Fig. 7), the U-net network accurately extracted the boundary information of a large range of glacial lakes based on the combination of 3, 5, and 6 bands and all bands of Landsat-8 in a complex mountain environment, and at the same time, removes the influence of mountain shadows [see Fig. 7(b) and (c)], which reflect the general's ability and time-space expansion of the U-net network. The 3, 5, and 6 bands of Landsat-8 image are sensitive to the water body, while the glacial lake is a special kind of water body, so it is necessary to extract glacial lake with the combination of 3, 5, and 6 bands. It is also found that the combination of 3, 5, and 6 bands can better extract the glacial lake. The self-attention U-net network can more accurately extract glacial lake boundary information [see Fig. 7(d) and (e)]. Compared with the U-net network, the range of glacial lake boundary extracted using the self-attention U-net network is larger and closer to the ground truth [see Fig. 7(f)]. However, in the segmentation result map, the U-net network cannot segment information well from a small glacial lake. There are many omissions, and the boundary segmentation is coarse (see Fig. 7(b) and (c), red ellipse), but the hillshade can be distinguished better (see Fig. 7). For small glacial lakes, the self-attention U-net network can extract the boundary more accurately. Especially when the input data are from all bands, the extraction result of the self-attention U-net network can be closer to the ground truth. This may be because the spectral characteristics of texture and geometry features for all bands are more abundant.

The extracted results were further analyzed to reveal the effectiveness of the new method. In order to compare the results of the extraction of glacial lakes more clearly, this study selected some typical areas and superimposed the extracted glacial lake boundary information on the Landsat-8 remote sensing images to reveal the accuracy at different scales (see Fig. 8). In Fig. 8, the first column [see Fig. 8(a)] is the original Landsat-8 image, the second column [see Fig. 8(b)] is the ground truth, the third column [see Fig. 8(c)] is the U-Net network segmentation result map for the 3, 5, and 6 bands combination, the fourth column [see Fig. 8(d)] is the U-net network segmentation result map for all bands, the fifth column [see Fig. 8(e)] is the self-attention U-net network segmentation result map for the combination of 3, 5, and 6 bands, and the sixth column [see Fig. 8(f)] is the self-attention U-net network segmentation result map for all bands.

As can be seen in Fig. 8, self-attention U-net network segmentation of the glacial lake boundary is more obvious, and the segmentation result similar to the segmentation ground truth can be obtained. The small glacial lake is segmented well and missed detections are relatively few. The influence of shadows can be removed at the same time, but surrounding features will be excessively extracted. In general, the self-attention U-net network can effectively combine the low-dimensional and high-dimensional feature information from an image. While digging the deep law of the image, it retains the low-dimensional feature information of the image and finds features that effectively distinguish the glacial lake from other ground objects.

As can be seen in Fig. 9, the training results of the self-attention U-net network in the binary graph of the training model have more abundant information (red circle) than that of the U-net network for the combination of 3, 5, and 6 bands and all bands, while the training results of all bands are better than the combination of 3, 5, and 6 bands. Because all bands have more information, and most of the bands are related to the features of the glacial lake, so the input of all bands into the neural network can assist the extraction of the spectral and texture features of the glacial lake. It can be seen from Fig. 9(e) and (f) that the manual labels were omitted. However, due to the redundant information of all bands and the fault tolerance of the network, some missing labels are also identified during the input of all bands. This shows that using all bands for the dataset can improve the classification accuracy.

B. Analysis of the Glacial Lake Extraction Performance

As can be seen from Fig. 10, the pixel accuracy of both the U-net and self-attention U-net network models in the initial stage of training is low. Both show high growth at the 30th epoch after which there is a little change. The U-net network model fluctuates and the self-attention U-net network model is relatively stable, indicating that the performance of the self-attention U-net network has been greatly improved. The U-net network and self-attention U-net network achieved about 80% classification accuracy in the training set of the combination of 3, 5, and 6 bands and all bands (11 bands), and 90% classification accuracy in the test set, respectively, but on the whole, the training process of the self-attention U-net network model is stable. The classification accuracy of the self-attention U-net network model in all bands was the highest. The loss rate during model training is very low. This suggests that the self-attention U-net network has good performance in glacial lake extraction.

In order to further quantitatively analyze the performance and extraction accuracy of the two models, the model performance was evaluated using the confusion matrix, Kappa coefficient, F1-score, MioU, and AUC. The results of these evaluation indicators are presented in Fig. 11. Using the U-net network, the true positive (TP) for the combination of 3, 5, and 6 bands and all bands was 59.45% and 72.90% [see Fig. 11(a) and (b)], respectively, while for the self-attention U-net network, the TP was 75.40% and 78.69% [see Fig. 11(c) and (d)], respectively.

For the combination of 3, 5, and 6 bands and all bands, the TP increased by 15.95% and 5.79%, respectively, compared with

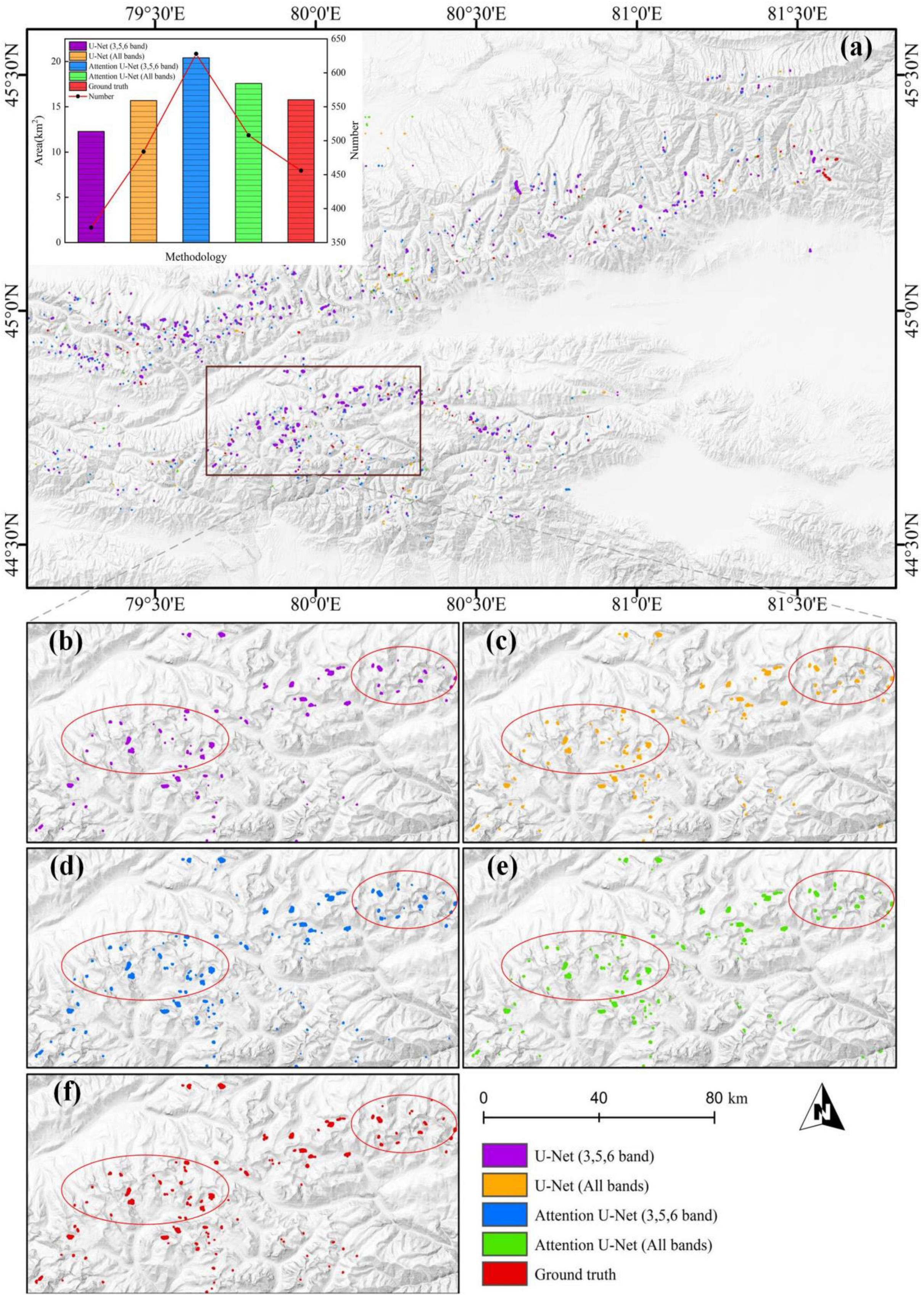


Fig. 7. Glacier lake extraction results based on U-net and attention U-net network.

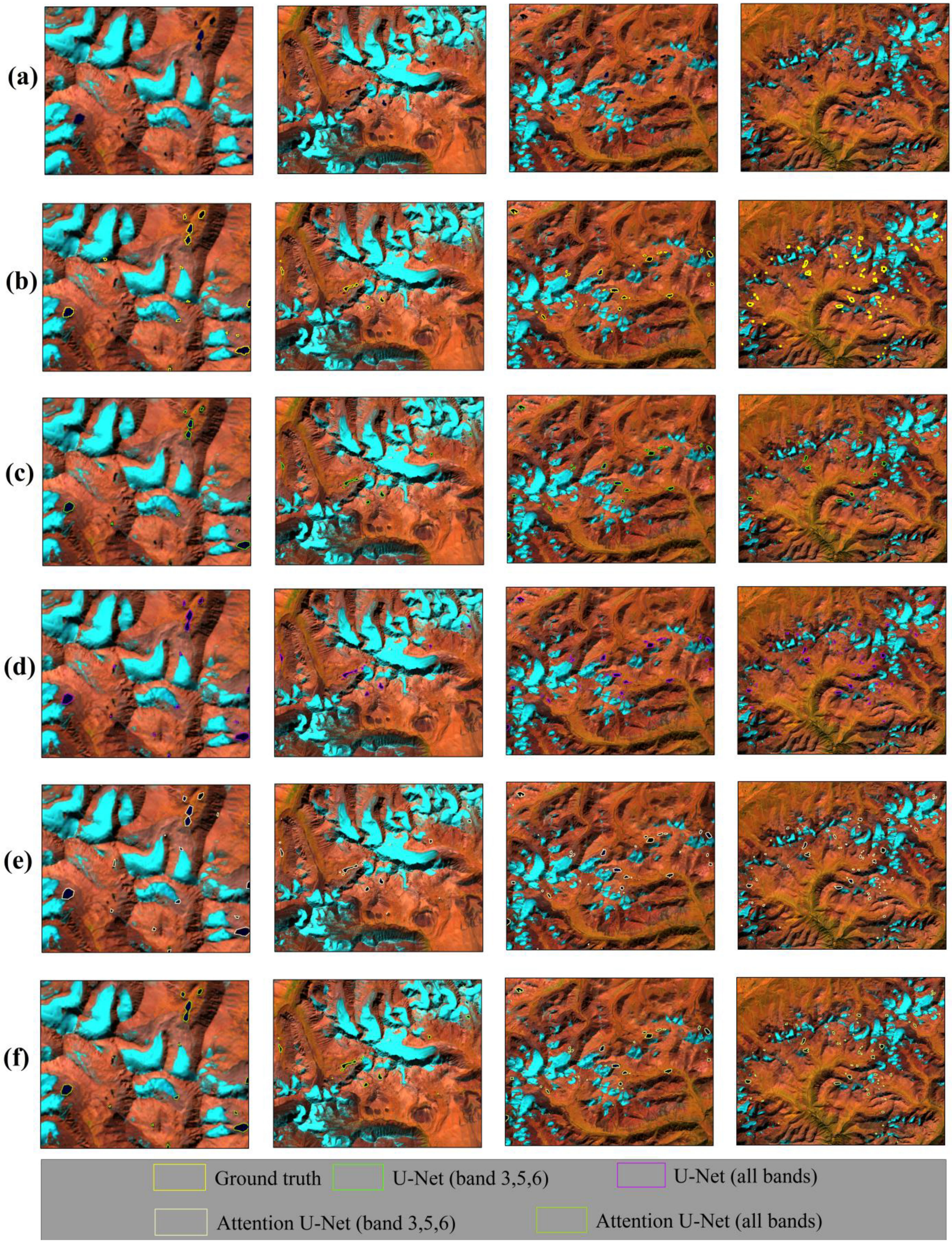


Fig. 8. Comparison of glacial lakes extraction results with different methods in typical areas of the study area. (a) Landsat-8 images. (b) Ground truth. (c) U-net (3, 5, and 6 bands). (d) Self-attention U-net (3, 5, and 6 bands). (e) U-net (all bands). (f) Self-attention U-net (all bands).

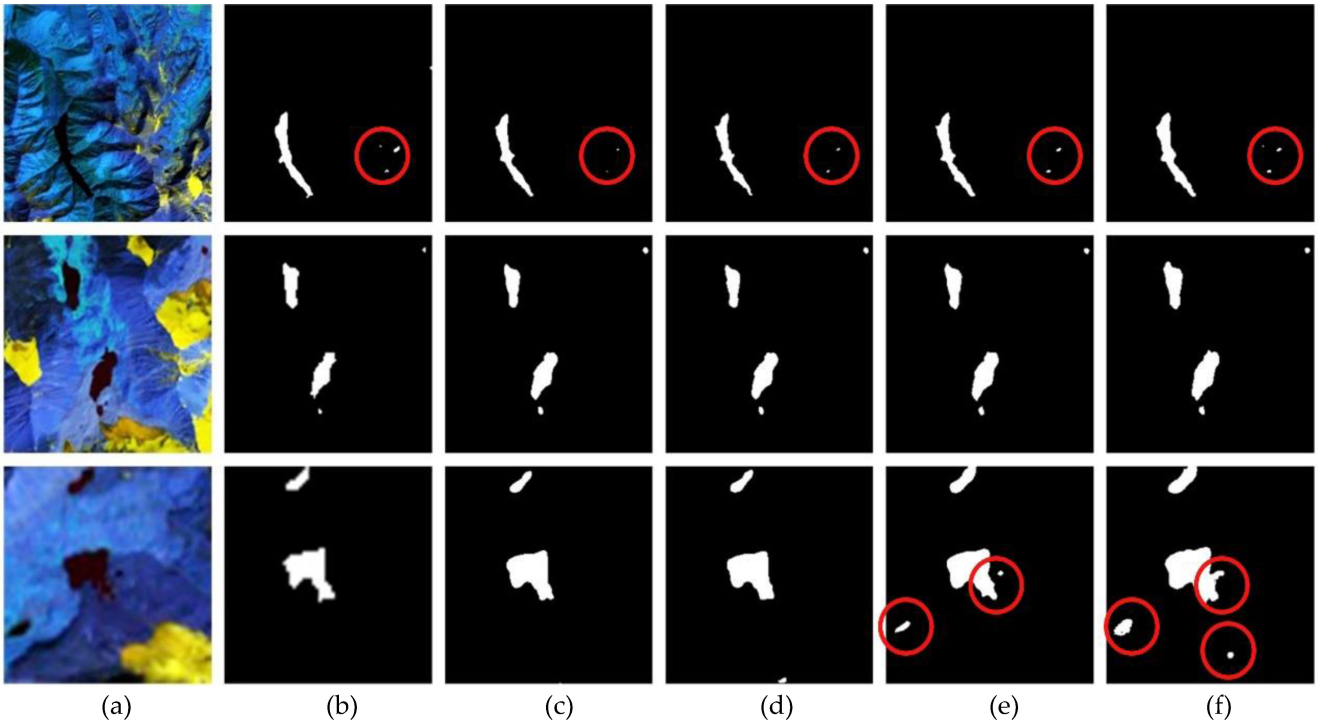


Fig. 9. Comparison of model training results. (a) Remote sensing images. (b) Ground truth. (c) U-net (3, 5, and 6 bands). (d) Self-attention U-net (3, 5, and 6 bands). (e) U-net (all bands). (f) Self-attention U-net (all bands).

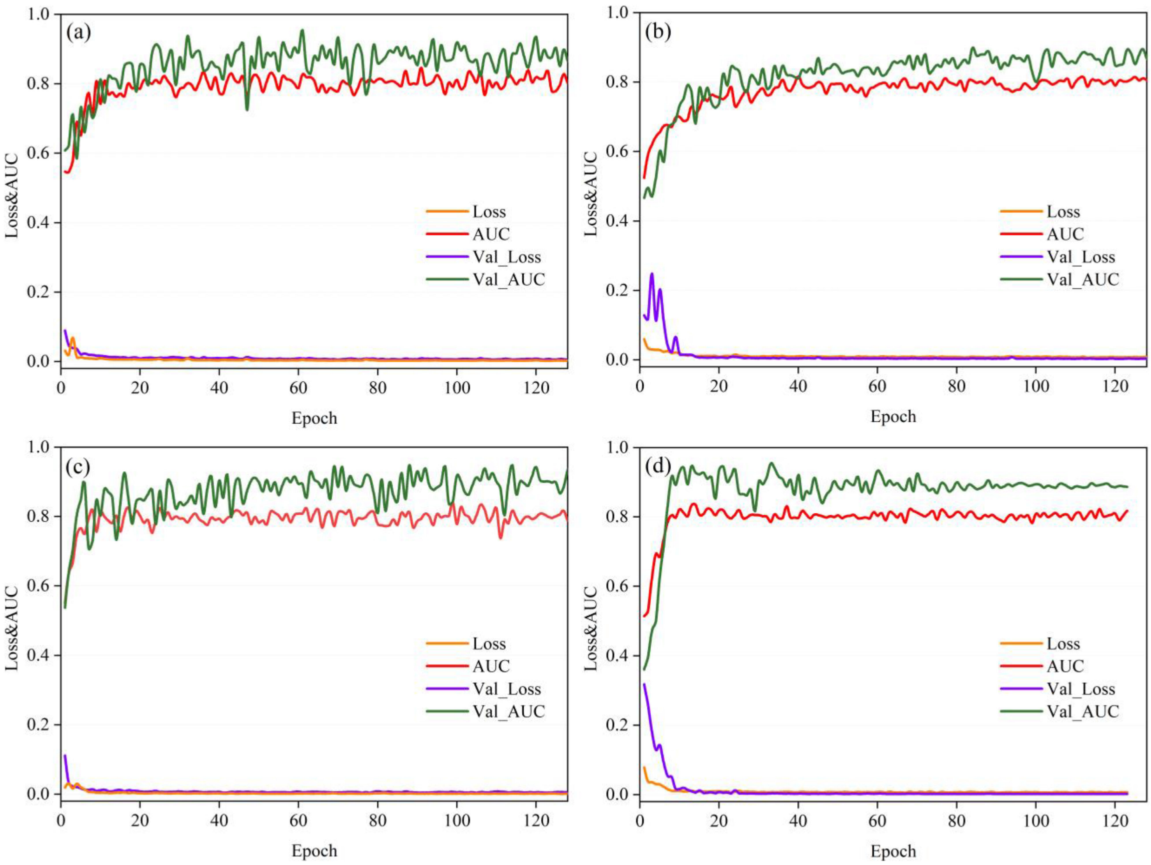


Fig. 10. U-net and attention U-net model training and testing curve. (a) U-net (3, 5, and 6 bands). (b) Attention U-net (3, 5, and 6 bands). (c) U-net (all bands). (d) Attention U-net (all bands).

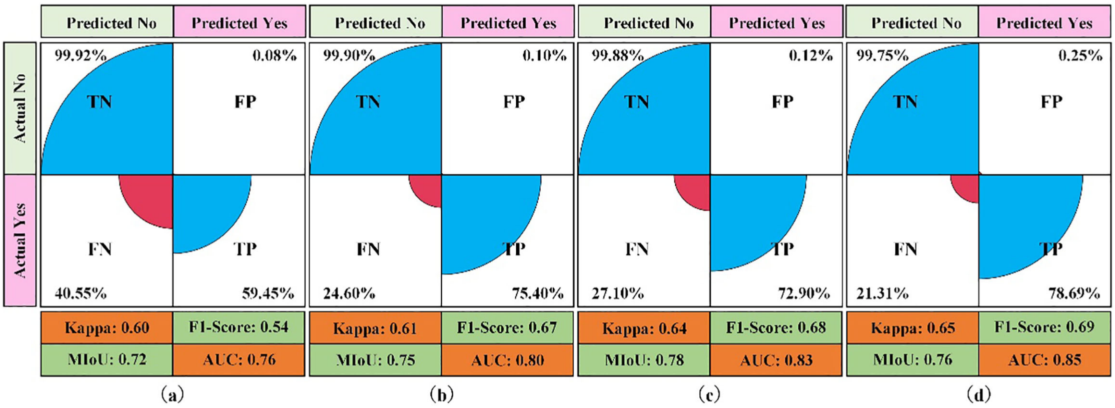


Fig. 11. Performance of extraction models trained on glacier lake inventories by confusion matrix, Kappa coefficient, F1-score, MIoU, and area under the ROC curve (AUC). (a) U-net (3, 5, and 6 bands). (b) Attention U-net (3, 5, and 6 bands). (c) U-net (all bands). (d) Attention U-net (all bands).

the U-net network. MIoU using the self-attention U-net network reached 0.75 and 0.76, respectively [see Fig. 11(b) and (d)]. In the entire large-scale research area, AUC for the combination of 3, 5, and 6 bands using the self-attention U-net network reached 80%, which is 4% higher than the U-net network model [see Fig. 11(a) and (b)]. For all bands, all index values of U-net and the self-attention U-net network model are higher than the combination of 3, 5, and 6 bands.

For the whole study area, the AUC based on the self-attention U-net network model can reach 0.85 [see Fig. 11(d)], suggesting that it is effective and feasible as a method for extracting glacial lakes and has good predictive performance. The self-attention U-net network model suppresses the weight of nonicy-lake features, slows down the impact of low image contrast on the model, suppresses the problem of pixel category imbalance, and improves the performance of the model, especially in its accuracy given all bands of data.

From the above analysis, it can be seen that the self-attention U-net network model can alleviate the problems of missed and false detections in low-contrast areas and small glacial lakes under complex backgrounds. The general's ability of the network model is relatively good. Therefore, the improved U-net network model can more accurately extract boundary information of glacial lakes in high mountains. The extraction method of glacial lakes in this study can provide a new technology for the rapid extraction of glacial lake disasters on the Sichuan–Tibet Railway.

V. Conclusion and Outlook

The proposed method applies remote sensing datasets for the accurate extraction of glacial lakes. At present, existing glacial lake extraction algorithms lack the ability to analyze glacial lake spectra and shape and texture features, and require manual design parameters to fine tune the automation of the algorithm. As a result, it cannot mine the depth features of glacier lakes in remote sensing images accurately enough. In the presented study, we introduced the attention mechanism into the step connection part of the U-net network to adjust feature weight, focus on learning glacial lake features, and strengthen the network to extract the glacial lake features. The attention U-net network model enhances the propagation of features, reduces information loss, strengthens the weight of the glacial lake areas, restrains the weight of irrelevant features, reduces the influence of low image contrast on the model, and deals with the variety of pixel categories in glacial lakes. These features improve the performance of the model. The effectiveness of this new method was proved by different evaluation indices. AUC for the whole study area reached 85.03% for all bands of the Landsat-8 images, which can, thus, meet the real-time demands of large-scale glacial lake disaster information acquisition. Thus, we can obtain a more reliable glacial lake extraction model. This study obtained a large-scale glacial lake dataset (2018) in the Alatau mountains of the Tianshan to provide data support for subsequent research on glacial lake disasters in the Tianshan mountains.

In follow-up work, more types of remote sensing images can be collected to effectively monitor the high mountain glacial lakes in a timely manner. In the training data, the number of samples under different background features should be increased to enhance the general's ability of the model. Therefore, the next step of the research work will focus on strengthening the extraction of glacial lakes under different background ground features and further improve the temporal and spatial scalability of the model.

Acknowledgment

The authors would like to thank the editors and anonymous reviewers for their valuable suggestions and the Geospatial Data Cloud for providing access to the orthorectified Landsat imagery.

References

[1] X. Yao, S. Liu, L. Han, M. Sun, and L. Zhao, "Definition and classification system of glacial lake for inventory and hazards study," *J. Geogr. Sci.*, vol. 28, no. 2, pp. 193–205, Feb. 2018.

[2] X. Wang *et al.*, "Changes of glacial lakes and implications in Tian Shan, central Asia, based on remote sensing data from 1990 to 2010," *Environ. Res. Lett.*, vol. 8, no. 4, Dec. 2013, Art. no. 044052.

[3] C. Prakash and R. Nagarajan, "Glacial lake inventory and evolution in northwestern Indian Himalaya," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5284–5294, Dec. 2017.

[4] Y. He, P. Dou, H. Yan, L. Zhang, and S. Yang, "Quantifying the main urban area expansion of Guangzhou using Landsat imagery," *Int. J. Remote Sens.*, vol. 39, no. 21, pp. 7693–7717, Jul. 2018.

[5] T. Che, L. Xiao, and Y.-A. Liou, “Changes in glaciers and glacial lakes and the identification of dangerous glacial lakes in the Pumqu river basin, Xizang (Tibet),” *Adv. Meteorol.*, vol. 2014, Jan. 2014, Art. no. 903709.

[6] G. Zhang, T. Yao, H. Xie, W. Wang, and W. Yang, “An inventory of glacial lakes in the third pole region and their changes in response to global warming,” *Global Planet. Change*, vol. 131, pp. 148–157, Aug. 2015.

[7] D. Li, D. Shangguan, and W. Huang, “Study on the area change of lakes Merzbacher in the Tianshan mountains during 1998-2017,” *J. Glaciol. Geocryol.*, vol. 41, no. 1, pp. 1–8, 2019.

[8] T. Bolch, M. Buchroithner, J. Peters, M. Baessler, and S. Bajracharya, “Identification of glacier motion and potentially dangerous glacial lakes in the Mt. Everest region/Nepal using spaceborne imagery,” *Natural Hazards Earth Syst. Sci.*, vol. 8, no. 6, pp. 1329–1340, Dec. 2008.

[9] T. Bolch, J. Peters, A. Yegorov, B. Pradhan, M. Buchroithner, and V. Blagoveshchensky, “Identification of potentially dangerous glacial lakes in the northern Tien Shan,” *Natural Hazards*, vol. 59, no. 3, pp. 1691–1714, Jun. 2011.

[10] W. Yang, M. Yang, and H. Qi, “Water body extracting from TM image based on BPNN,” *J. Sci. Surv. Mapping*, vol. 37, no. 1, pp. 148–150, Jul. 2012.

[11] J. Li, T. A. Warner, Y. Wang, J. Bai, and A. Bao, “Mapping glacial lakes partially obscured by mountain shadows for time series and regional mapping applications,” *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 615–641, 2019.

[12] P. Dou and Y. Chen, “Remote sensing imagery classification using AdaBoost with a weight vector (WV AdaBoost),” *Remote Sens. Lett.*, vol. 8, no. 8, pp. 733–742, Apr. 2017.

[13] P. Du *et al.*, “Advances of four machine learning methods for spatial data handling: A review,” *J. Geovis. Spatial Anal.*, vol. 4, no. 13, May 2020, Art. no. 13.

[14] J. C. Luo, Y. Sheng, Z. Shen, J. L. Li, and L. Gao, “Automatic and high-precise extraction for water information from multispectral images with the step-by-step iterative transformation mechanism,” *J. Remote Sens.*, vol. 13, no. 4, pp. 610–615, 2009.

[15] H. Zhao, “The research of glacial lake extraction based on Landsat8 OLI imagery in high mountain region of Asian,” *Inst. Remote Sens. Digit. Earth, Chin. Acad. Sci.*, 2018.

[16] L. Zhu *et al.*, “Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network,” *Sensors*, vol. 20, no. 6, Mar. 2020, Art. no. 1576.

[17] C. Pelletier, G. Webb, and F. Petitjean, “Temporal convolutional neural network for the classification of satellite image time series,” *Remote Sens.*, vol. 11, no. 5, Mar. 2019, Art. no. 523.

[18] X. X. Zhu *et al.*, “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[19] L. Zhong, L. Hu, and H. Zhou, “Deep learning based multi-temporal crop classification,” *Remote Sens. Environ.*, vol. 221, pp. 430–443, Feb. 2019.

[20] D. H. T. Minh *et al.*, “Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR sentinel-1,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 464–468, Mar. 2018.

[21] V. Badrinarayanan, A. Handa, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling,” CVPR, 2015.

[22] Y. Wang *et al.*, “Extracting urban water by combining deep learning and google earth engine,” *Comput. Sci*., 2019, *arXiv: 1912.10726*.

[23] P. Dou, H. Shen, Z. Li, X. Guan, and W. Huang, “Remote sensing image classification using deep–shallow learning,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3070–3083, Mar. 2021.

[24] P. Dou and C. Zeng, “Hyperspectra l image classification using feature relations map learning,” *Remote Sens.*, vol. 12, no. 18, Sep. 2020, Art. no. 2956.

[25] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 3431–3440.

[26] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.

[27] G. Wu, Q. Chen, S. Ryosuke, Z. Guo, X. Shao, and Y. Xu, “High precision building detection from aerial imagery using a U-net like convolutional architecture,” *Acta Geodaetica et Cartographica Sinica*, vol. 47, no. 6, pp. 864–872, 2018.

[28] J. M. Su, L. X. Yang, and W. P. Jing, “A U-net based semantic segmentation method for high resolution remote sensing image,” *Comput. Eng. Appl.*, vol. 55, no. 7, pp. 207–213, 2019.

[29] H. He *et al.*, “Water body extraction of high resolution remote sensing image based on improved U-net network,” *Int. J. Geo-Inf. Sci.*, vol. 22, no. 10, pp. 2010–2022, 2020.

[30] N. Wang *et al.*, “Application of U-net model to water extraction with high resolution remote sensing data,” *J. Remote Sens. Land Resour.*, vol. 32, no. 1, pp. 35–42, 2020.

[31] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[32] G. Fang, “Automatic extraction of water body information based on Landsat 8 satellite OLI image,” *Chin. J. Soil Sci.*, vol. 46, no. 6, pp. 1284–1288, 2015.

[33] H. Chang *et al.*, “Research on tunnel crack segmentation algorithm based on improved U-net network,” *Comput. Eng. Appl. J.*, pp. 1–11, Nov. 2020.

[34] Z. Liu, Y. Cao, Y. Wang, and W. Wang, “Computer vision-based concrete crack detection using U-net fully convolutional networks,” *Autom. Construction*, vol. 104, pp. 129–139, Aug. 2019.

[35] O. Ozan *et al.*, “Attention U-net: Learning where to look for the pancreas,” MIDL, 2018.

[36] A. Vaswani *et al.*, “Attention is all you need,” NIPS, 2017, *arXiv: 1706.03762*.

[37] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[38] X. Zhang *et al.*, “Survey of the deep learning models for image semantic segmentation,” *J. Chin. High Technol. Lett.*, vol. 27, no. Z1, pp. 808–815, Jul. 2017.

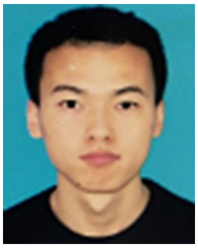
[39] B. Yang, “Change detection of high resolution remote sensing image based on deep learning,” *China Univ. Mining Technol.*, 2019.

Yi He received the B.S. degree in geographic information system from Lanzhou Jiaotong University, Lanzhou, China, in 2011, and the Ph.D. degree in earth system science from Lanzhou University, Lanzhou, in 2016.

He has been a Postdoctoral Researcher with the School of Environment and Municipal Engineering, Lanzhou Jiaotong University, Lanzhou, China. He is currently an Associate Professor with the Faculty of Geomatics, Lanzhou Jiaotong University. His research interests include disaster remote sensing, ecological remote sensing, image processing, and time-series InSAR prediction based on deep learning.

Sheng Yao received the B.E. degree in digital media technology from the North China University of Technology, Beijing, China, in 2020. He is currently working toward the M.S. degree in surveying and mapping with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China.

His research interest focuses on the application of remote sensing in deep learning.



Wang Yang received the B.E. degree in remote sensing science and technology in 2019 from Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in surveying and mapping with the Faculty of Geomatics.

His research interests include InSAR data processing technology, machine learning, and remote sensing image information extraction.

Haowen Yan received the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2002, and the Ph.D. degree in waterloo from Waterloo University, Waterloo, ON, Canada, in 2014.

He is an Editor-in-Chief of *the Journal of Geovisualization and Spatial Analysis*. He is the Dean of the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China, and the Director of the National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou. His research interests include WeMaps, automated map generalization, spatial relations, geovisualization and spatial analysis, spatial data fingerprinting and watermarking, history of maps, and cartography.

Lifeng Zhang received the M.S. degree in cartography and geographic information system and the Ph.D. degree in environmental science and engineering both from Lanzhou Jiaotong University, Lanzhou, China, in 2010 and 2017, respectively.

He is an Associate Professor with Lanzhou Jiaotong University. His research interests include ecological environment remote sensing monitoring, visualized analysis, and analysis of land use change.

Zhiqing Wen received the B.E. degree in surveying and mapping engineering from Chang'an University, Xian, China, in 2020. He is currently working toward the M.S. degree in surveying and mapping with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China.

His research interests include remote sensing image processing and remote sensing application.

Yali Zhang received the B.E. degree in remote sensing science and technology in 2020 from Lanzhou Jiaotong University, Lanzhou, China, where she is currently working toward the M.S. degree in surveying and mapping with the Faculty of Geomatics.

Her research interests include InSAR data processing technology, extraction of glacier movement velocity, and remote sensing image fusion.

Tao Liu received the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2011.

He is currently a Professor with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China. He is a member of the Theory and Method Work Committee of China Geographic Information Industry Association. His research interests include spatial relation theory, GIS and RS integration, and application and development of GIS and RS.